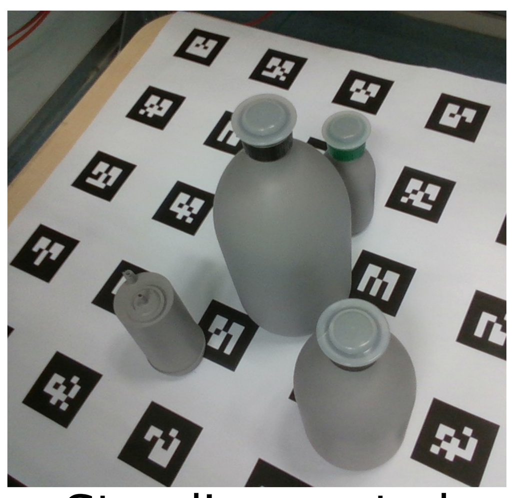
HSLU Hochschule Luzern

Technik & Architektur

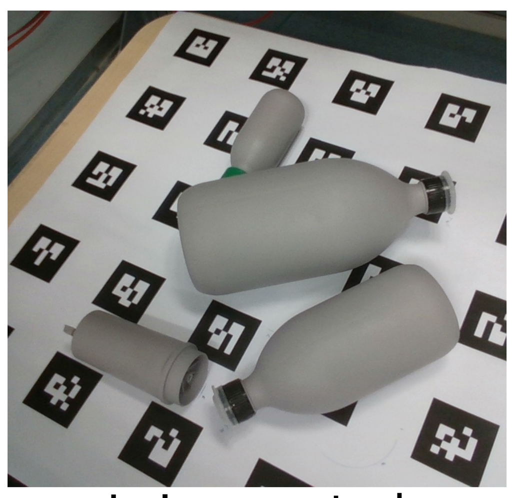
Master Thesis - Mechatronics & Automation

Neural Radiance Fields-enabled Transparent Object Pose Estimation

Captured Scenes

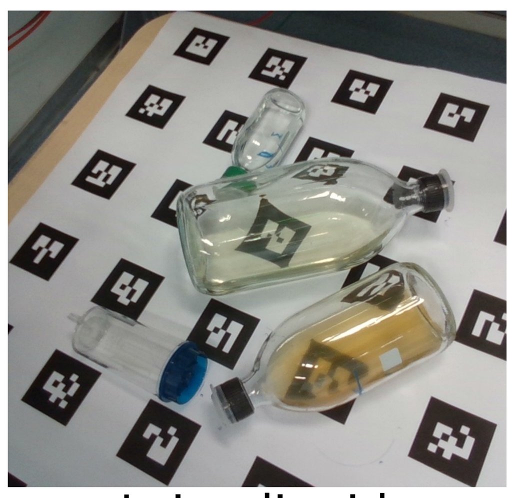


Standing coated

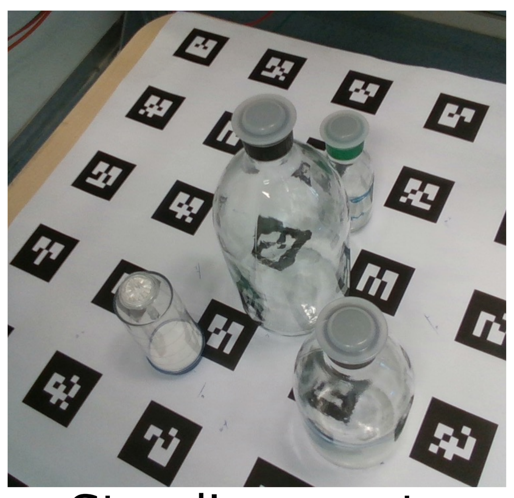


Lying coated

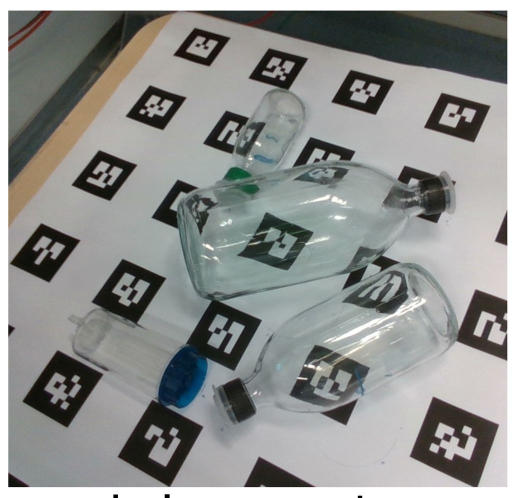
Standing liquid



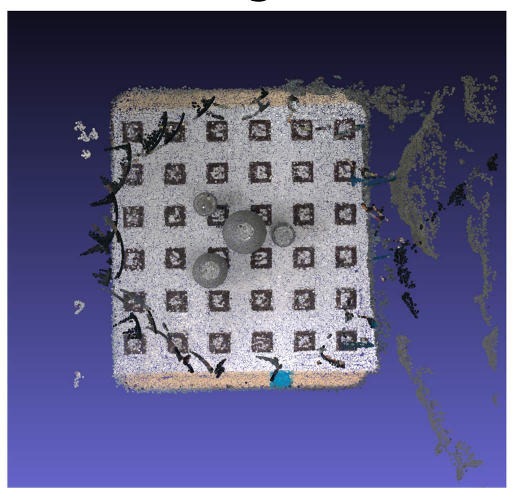
Lying liquid

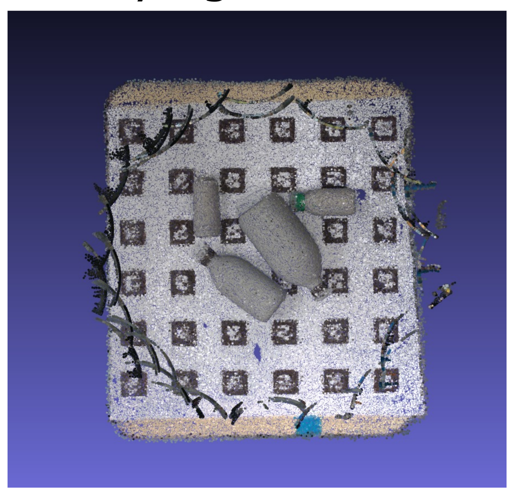


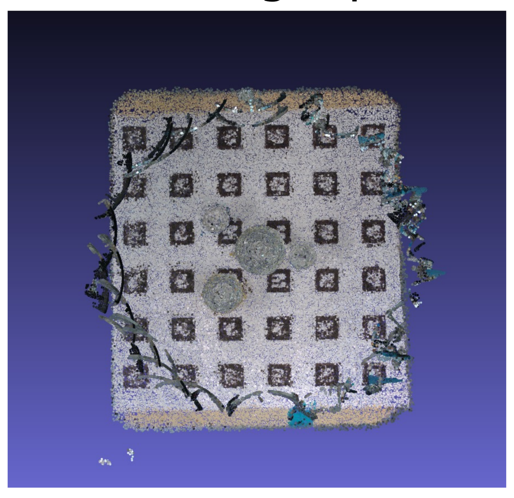
Standing empty

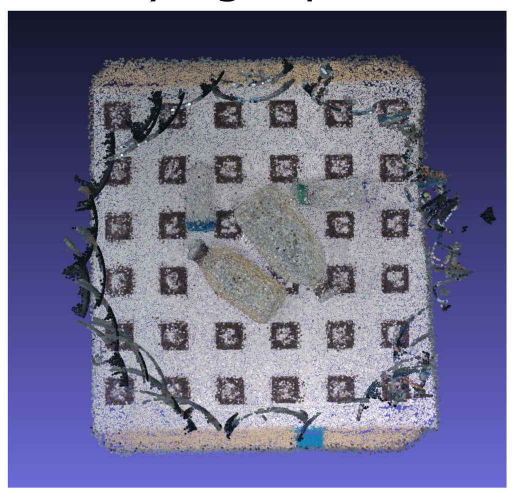


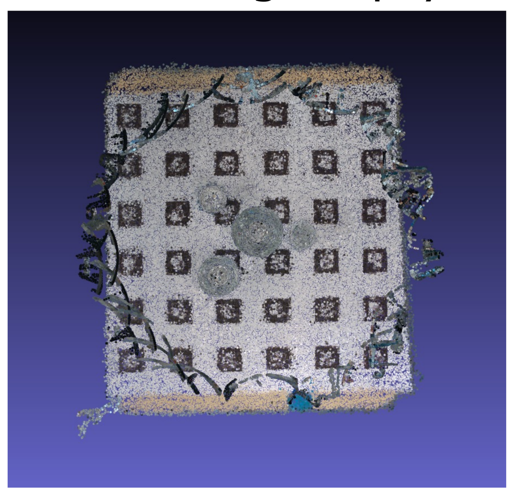
Lying empty

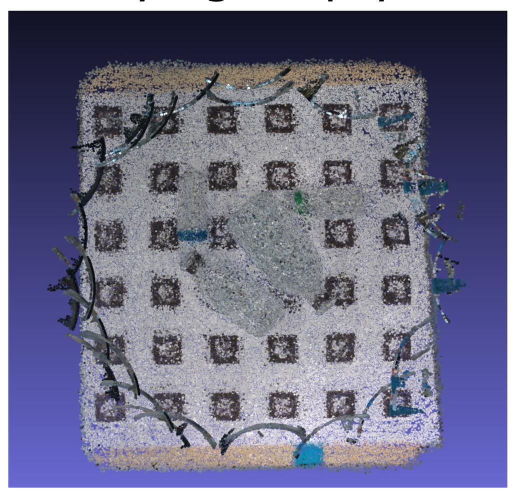




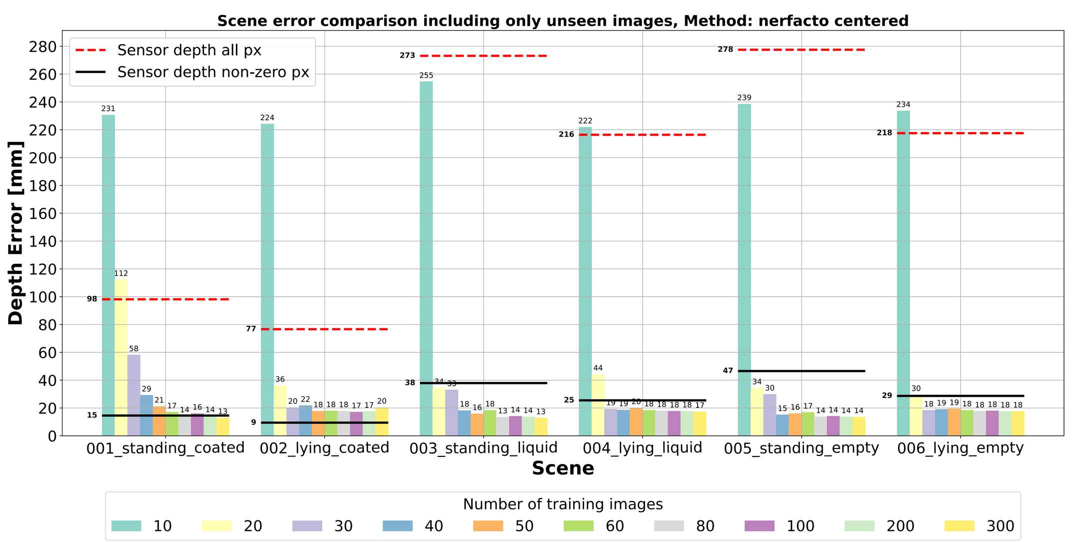




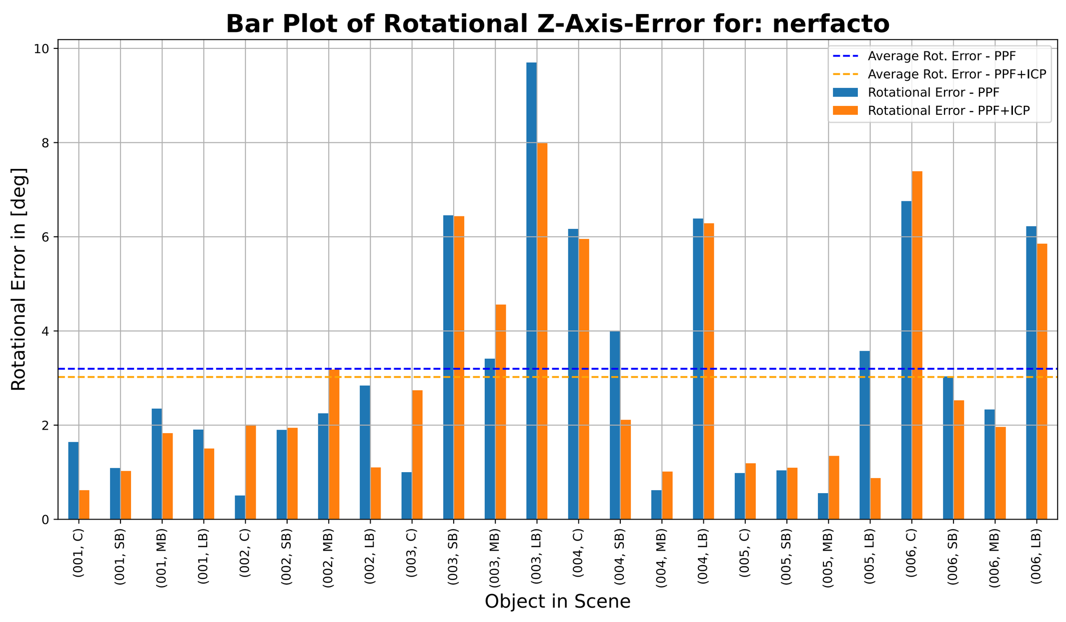




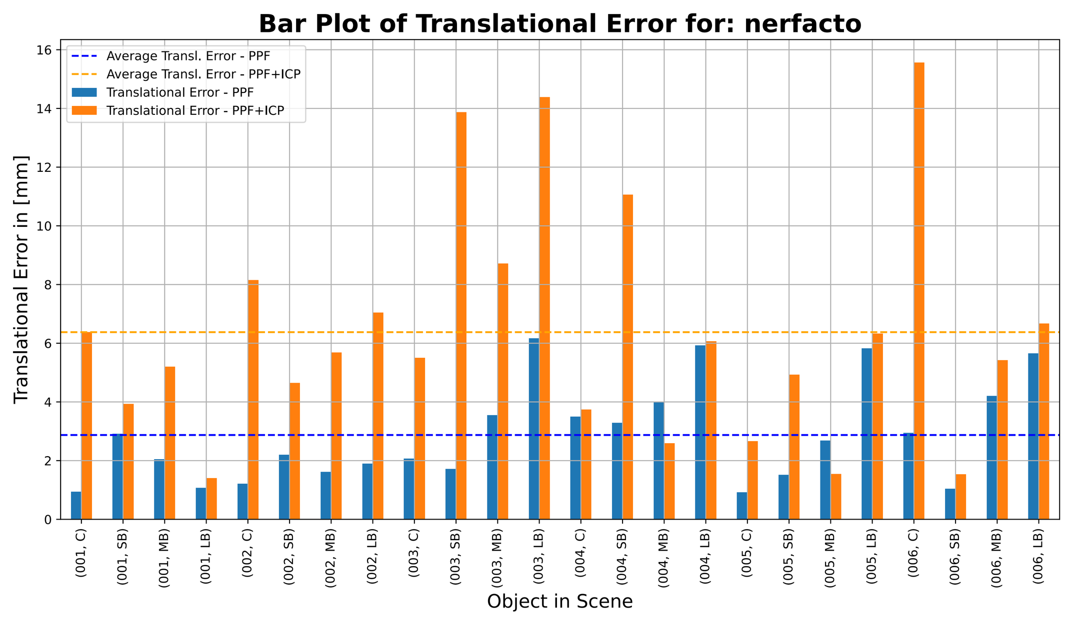
Recovered point clouds from Nerfacto method



Depth error results for all six scenes



Rotational error results for each object



Translational error results for each object

Motivation and Problem Statement

Transparent objects are ubiquitous in our lives but lack discernible visual features and surface details, which is problematic for pose estimation methods that heavily rely on texture or depth maps. This thesis evaluated the performance of Neural Radiance Fields (Mildenhall et al., 2020) to reconstruct valid depth maps of scenes containing transparent objects, dependent on the number of used training images.

Experiments

The full set of images contained 424 images for each of the six scenes. In order to evaluate the number of necessary training viewpoints, ten subsets using 10, 20, 30, 40, 50, 60, 80, 100, 200, and 300 images for NeRF training were generated. To ensure good coverage of the whole scene in each subsample, the farthest point sampling (FPS) strategy was chosen. For both NeRF methods, the hyperparameters sigma and iterations were set at 10 and 20.000 respectively. The axis-aligned bounding box (aabb) parameter was set for instant-NGP at 4 and for Nerfacto at 1.

Results

Depth Error

Fifty training images were enough for the Nerfacto method to reconstruct a more accurate depth scene than the depth sensor could. Instant-NGP proved to outperform the depth sensor when trained with at least 200 images. Both methods generalized well to the 124 unseen test images.

Translational Pose Estimation

Nerfacto's results were highly accurate, with an average translational error for all objects and all scenes for the PPF-only and the ICP-refined estimates below 6.37 mm.

Instant-NGP's average translational error was much higher than Nerfacto's (82mm) due to the sparsely reconstructed point cloud, which led PPF to estimate some objects' poses within another object.

The sensor's results had the highest average translational error (116mm) due to suffering from the same "one object in another object" problem.

Rotational Pose Estimation

Nerfacto's results showed close alignment for the ground truth and the estimated z-axes for most of the objects in most scenes.

Instant-NGP's rotational error results suffered from severe rotational misalignment, even for the coated scenes.

The sensor's rotational errors were high for almost all objects in all scenes.

As with the translational error, the results for the Nerfacto method were accurate, while the other two methods failed to enable PPF to provide a reliable rotational estimate. This was again owed to the sparse and inaccurate point cloud.

Julius Roschmann

Advisor:
Prof. Dr. Björn Jensen

Expert:
Thomas Estier

Partner:
Vienna University of Technology – Vision for Robotics Group

FH Zentralschweiz

MSE | MASTER OF SCIENCE IN ENGINEERING